A latitudinal gradient of reference genomes

Ethan B. Linck[1]*, Carlos Daniel Cadena[2]

[1]*Department of Ecology, Montana State University*

[2]*Departamento de Ciencias Biológicas, Universidad de los Andes*

*Correspondence: ethan.linck@montana.edu

Abstract: Global inequality rooted in legacies of colonialism and uneven development can lead to systematic biases in scientific knowledge. In ecology and evolutionary biology, findings, funding and research effort are disproportionately concentrated at high latitudes while biological diversity is concentrated at low latitudes. This discrepancy may have a particular influence in fields like phylogeography, molecular ecology and conservation genetics, where the rise of genomics has increased the cost and technical expertise required to apply state-of-the-art methods. Here we ask whether a fundamental biogeographic pattern—the latitudinal gradient of species richness in tetrapods—is reflected in available reference genomes, an important data resource for various applications of molecular tools for biodiversity research and conservation. We also ask whether sequencing approaches differ between the Global South and Global North, reviewing the last five years of conservation genetics research in four leading journals. We find that extant reference genomes are scarce relative to species richness at low latitudes, and that reduced-representation and whole-genome sequencing are disproportionately applied to taxa in the Global North. We conclude with recommendations to close this gap and improve international collaborations in biodiversity genomics.

Keywords: biodiversity genomics, conservation genetics, tetrapods, macroecology

Introduction: Scientific knowledge reflects both reality and the economic and social conditions that shape the practice of research (Hull 1990; Logino 2002; Stephan 2012). The coincidence of the Scientific Revolution and imperial expansion in Western Europe in the 16th through 18th centuries led to the accumulation and concentration of global capital in institutions of higher learning in the United Kingdom, France, Spain, the Netherlands, and Germany, funding foundational investigations in the natural sciences. Industrialization and the rise of capitalism in the 19th century stoked the boiler of the research endeavor, permitting the continued expansion of universities and birthing a learned aristocratic class with disposable time and income to commit to natural history (Opitz 2004; Opitz 2006). This legacy persists to this day: Research expenditures concentrated in a handful of majority Anglophone countries (May 1997), which in turn produce a disproportionate amount of scholarship (May 1997; King 2004) and often drive citation networks (Pasterkamp et al. 2007; Meneghini et al. 2008).

Ecology and evolutionary biology (EEB) face a unique challenge in light of global scientific inequality, as while funding and research effort are disproportionately concentrated at high latitudes (Melles et al. 2019), biological diversity is concentrated at low latitudes (Willig et al. 2003). The consequences of this discrepancy are many, including systematic biases in research effort (Titley et al. 2017) and taxonomy (Freeman & Pennell 2021), major gaps in our understanding of natural history and species distributions (Collen et al. 2008; Feely & Sillman 2011), and a blinkered view of patterns of diversity and diversification across the tree of life (Reddy 2014; Cornwell et al. 2019). Increased research attention on tropical ecosystems has therefore been a stated priority for decades (Collen et al. 2008; Nori et al. 2020), but despite some promising strides towards international collaborations (Perez et al. 2018), partners in the Global North frequently set research agendas (Bradeley 2008; Asase et al. 2022), leading to

accusations of parachute science and persistent gaps in data quantity, quality, and type (Stocks et al. 2008; Soares et al. 2023).

EEB disciplines where data collection and analysis are equipment- and cost-intensive are especially likely to show increased discrepancies between the Global North and Global South. In the 1960s and 70s, assays of molecular genetic markers in non-model organisms laid the groundwork for phylogenetic systematics to shift away from a traditional focus on morphology, increasing statistical power and requiring increasingly expensive tools and specialized skills (Hillis et al. 1996). Concurrently, the rise of phylogeography brought an evolutionary perspective to population biology, altering views on the fundamental units of management and conservation (Avise 2000). In the late 2000s and early 2010s, reduced representation high-throughput sequencing using library preparation methods such as RADseq and target-capture began to proliferate as an alternative to traditional Sanger sequencing (Eklbom et al. 2011; Lemmon & Lemmon 2013), boasting the advantages of rapidly generating much larger datasets and requiring little to no prior information about a focal taxon's genome.

A third revolution is currently underway, as low-coverage whole-genome resequencing (hereafter WGS) begins to supplant reduced representation methods in some fields and taxa (Ellegren et al. 2014; Toews et al. 2016). Unlike RADseq, target capture, and related approaches, WGS typically requires a high-coverage reference genome with which to align samples. Because sequencing an individual at sufficient depth remains costly—and because assembling such large data is computationally intensive—generating high-quality novel reference genomes remains difficult for many labs and, by extension, species.

The problem may be particularly acute in conservation biology where decades of concern about the so-called "conservation genetics gap" have highlighted a persistent disconnect between

70 the importance managers place on genetic information and its actual use in decision making
71 (Taylor et al. 2017). Hypotheses to explain the conservation genetics gap include skepticism
72 about its importance, the specialized knowledge required for analysis and interpretation, and cost
73 (Hoban et al. 2013), though some recent discussion suggests the term itself obfuscates by
74 lumping the diverse set of obstacles (or "spaces") between research and implementation (i.e.,
75 Toomey et al. 2017; see solutions reviewed in Hogg et al. 2024). At low-resource institutions in
76 the tropics a broader shift towards WGS approaches may further discourage the development of
77 the field in the very regions where the greatest number of critically endangered species are found
78 (Vamosi & Vamosi 2008; Bertola et al. 2024), especially if technological advances become
79 publication requirements at high impact and broadly read journals.
80 Here we ask whether a fundamental biogeographic pattern—the latitudinal gradient of
81 species richness in tetrapods—is reflected in reference genomes available through the US
82 National Institute of Health's National Center for Biotechnology Information Genome Browser,
83 an important data resource for various applications of molecular tools for biodiversity research
84 and conservation. We hypothesized that NCBI contributor affiliations combined with inequities
85 in economic development and access to scientific resources would lead the number of species
86 with assembled genomes to be greatest in the temperate zone, not the tropics. We also ask
87 whether sequencing approaches differ between the Global South and Global North, reviewing
88 the last five years of conservation genetics research in four leading English-language journals.
89 We conclude by discussing strategies to improve international collaborations in biodiversity
90 genomics and boost the representation of species from the Global South in sequence databases.
91
92 **Methods:**

93 *Reference genomes, species richness, and latitude.* We used the National Center for
94 Biotechnology Information (NCBI) Datasets command-line tools v.16.19.0 (O'Leary et al. 2024)
95 to download taxonomy metadata for the subset of species with an assembled reference genome in
96 the following taxa: birds (Class: Aves), mammals (Class: Mammalia), squamates (Order:
97 Squamata), amphibians (Class: Amphibia), turtles (Order: Testudines), crocodilians (Order:
98 Crocodilia) and tuataras (Order: Rhynchocephalia). We selected these groups—together
99 comprising extant tetrapods—to provide a snapshot of animal diversity in relatively well-studied
100 clades with different ecologies and evolutionary histories, while restricting the total dataset to a
101 computationally manageable size. From this initial list we retained species with an exact match
102 to the Global Biodiversity Information Facility's (GBIF) Backbone Taxonomy using rgbif
103 v.3.8.0 (Chamberlain et al. 2024) and downloaded all observations of each backed by
104 georeferenced voucher specimens in natural history museum collections (NHCs), excluding
105 those without coordinates and those flagged for geospatial issues (n= 3,006,946). We repeated
106 this process for all species in each higher-level taxon represented in our list of reference genomes
107 (i.e., downloaded metadata for all georeferenced tetrapod specimens on GBIF; n= 9,303,258).
108 DOIs for each download are available in the References section below (GBIF.org 2024a;
109 GBIF.org 2024b; GBIF.org 2024c; GBIF.org 2024d; GBIF.org 2024e; GBIF.org 2024f;
110 GBIF.org 2024g; GBIF.org 2024h).
111 Filtering these aggregated datasets to contain only species with 10 or more specimen
112 records, we generated convex hull polygons for each as a coarse approximation of their
113 geographic distribution using the R package sf v.1.0-16 (Pebesma 2018; Pebesma & Bivand
114 2023). Overlaying these on a shapefile of Earth's landmasses from rnaturalearth v.1.0.1
115 (Massicotte & South 2024), we calculated species richness as the number of overlapping convex

116 hulls in 2-degree x 2-degree grid cells, statistically standardizing this value by subtracting
117 observed mean global species richness and dividing by its standard deviation. We subtracted the
118 number of species with reference genomes from total species richness to determine the regions
119 with the largest representation gap in genomic resources, again standardizing the difference. To
120 assess the significance and slope of a correlation between species richness and the absolute value
121 (or modulus) of latitude in decimal degrees, we performed simple linear regressions in R v.4.4.0
122 (R Core Team 2024), analyzing species with reference genomes and our full dataset separately.
123
124 *Literature review.* To evaluate how the geography of authorship might impact sequencing
125 strategy of studies in conservation biology, we performed a restricted Web of Science literature
126 search on 29 June 2024 for English-language conservation genetics papers published in the last
127 five years in the journals *Conservation Genetics*, *Molecular Ecology*, *Journal of Heredity*, and
128 *Conservation Biology,* selected for frequently publishing empirical work on non-model
129 organisms. We used the queries 'SO="Conservation Genetics"' and 'SO=("Molecular Ecology"
130 OR "Journal of Heredity" OR "Conservation Biology") AND (TS="Conservation Genet*" OR
131 KP="Conservation Genet*" OR TI="Conservation Genet*"'), excluding reviews, genome
132 announcements, meta-analyses, preprints, and studies that were purely simulations. Our criteria
133 aimed to achieve a tractable sample size for careful study (<1000 papers) while covering the
134 period in which WGS became commonly used for the conservation genetics of non-model
135 organisms (Fuentes-Pardo & Ruzzante 2017; Hohenlohe et al. 2021).
136 We then manually reviewed each study, first assigning the home institution of its first and
137 last author to the Global North, Global South or both (i.e., joint affiliations) using the 2024 UN
138 Trade and Development Classifications. Because the number of middle authors varied widely

139 across our sample, we assessed their affiliations on a binary basis, indicating only whether a
140 contributor from an institution from the Global South was present outside of the lead and senior
141 positions. Synthesizing these data, we assigned papers to mutually exclusive groups based on
142 whether they included one or more Global South authors or only Global North authors. Next, we
143 categorized each study's sequencing approach as reduced representation, WGS, Sanger
144 sequencing, microsatellites, or other, and described its overall focus using tiered categories based
145 on discussion in Bertola et al. 2024. These tiers were: 1) Taxonomy / systematics, identification,
146 or sexing; 2) Phylogeography / population genetic structure, estimating genetic diversity, and
147 inferring demographic history; and 3) Detecting outlier loci, quantifying runs of homozygosity,
148 and evaluating adaptive potential. When studies employed more than one sequencing approach
149 or addressed goals belonging to multiple tiers, we assigned them to a single category based on
150 their most data-intensive method or question. To explore geographic patterns in sequencing
151 effort, we assessed whether each study's taxonomic sampling included 1) at least one species
152 distributed in the Global South and 2) at least one species distributed in the Global North.
153 Because some studies included multiple taxa and some species are broadly distributed or migrate
154 between regions, these categories were not mutually exclusive. To evaluate whether geographic
155 representation in conservation genetics changed over the period covered by our review, we
156 performed logistic regression using the stats package R v.4.4.0, treating the presence or absence
157 of an author from the Global South as a binary outcome variable and year as the sole
158 independent variable.
159
160 **Results:**

161 *Sampled Gradients of Species Richness.* Our list of tetrapod reference genomes from
162 NCBI included 1159 bird species, 795 mammal species, 123 amphibian species, 39 turtle
163 species, 6 crocodile species, and the tuatara (*Sphenodon punctuatus*). This total represented 6.8%
164 of the 30,832 tetrapod species with a georeferenced preserved specimen record on GBIF. Species
165 with a reference genome were associated with 3,048,136 specimens, or 40.7% of the 7,478,867
166 tetrapod specimens with available metadata. Following filtering out species with 10 or fewer
167 observations, we retained 1,859 species with a reference genome, or 8.6% of the 21,583 tetrapod
168 species meeting the same criteria.
169 Minimum tetrapod species richness in both datasets was 1. Maximum species richness
170 calculated from the reference genome dataset was 705, occurring in a grid cell centered on south
171 Florida, USA (**Figure 1C**). Maximum species richness in the full dataset was 2698, occurring in
172 a grid cell centered on the western Amazon basin and east slope of the Andes in Ecuador near the
173 Peruvian border (**Figure 1C**); this was also where the greatest gap between sequenced species
174 and total species richness was observed (**Figure 1A**).
175 Species richness was negatively related to the absolute value latitude in both regressions,
176 albeit with a much steeper slope when data from all tetrapod species were included (reference
177 genomes only, $\beta = -4.892$, adjusted $R^2=0.6439$; $p<0.001$; full data, $\beta = -18.65$, adjusted
178 $R^2=0.7856$ $p<0.001$) (**Figure 1B**). Because data visualization indicated there might be a distinct
179 breakpoint in the relationship at mid latitudes—potentially reflecting a transition from the
180 influence of sampling effort effects to true biogeographic signal—we fit an additional piecewise
181 linear regression model with the R package segmented v.2.1-0. This model identified a
182 breakpoint at 39.819, fitting a segment with a slope of $\beta = -0.312$ before it, and a segment with a
183 slope of $\beta = -7.3634$ after it ($p=0.0408$; adjusted $R^2=0.7184$).

184

185 *Literature Review.* After excluding papers that did not meet our stated criteria, we
186 reviewed 394 empirical conservation genetics articles published between 1 January 2020 and 29
187 June 2024. This list included 345 papers from the journal *Conservation Genetics,* 11 papers from
188 *Journal of Heredity*, 35 articles from *Molecular Ecology,* and 3 articles from *Conservation*
189 *Biology*. Of these, 62 included a first or senior author from the Global South; 40 more included
190 Global South scientists as middle authors. Nearly 87% (n=342) of papers had a first or senior
191 author from the Global North. Ninety-eight included sampling from a focal taxon or focal taxa in
192 the Global South, with 277 sampling a taxon or taxa from the Global North. From 2020 to 2024,
193 the odds of a paper including an author from the Global South did not significantly increase (log-
194 odds 1.1096; 95% CI -0.5864-0.2667; p=0.210). Microsatellites were the most used sequencing
195 strategy among Global South authors and in studies of Global South taxa, while reduced
196 representation genome sequencing approaches were most common in the Global North. For
197 authors and taxa in both the Global South and the Global North, Tier 2 studies (phylogeography /
198 population genetic structure, estimating genetic diversity, or inferring demographic history) were
199 most common. Further details are provided in **Table 1** and **Table 2**.

200

201 **Discussion:** Reference genomes available through the US National Institute of Health's
202 National Center for Biotechnology Genome Browser fail to reflect the overwhelming
203 concentration of tetrapod species richness in the tropics and are strongly biased towards species
204 at mid-latitudes in the Northern Hemisphere (**Figure 1**). This pattern is almost certainly a result
205 of global inequalities in economic development and its resulting effects on research productivity
206 (May 1997; King 2004). Its consequences will likely include increasing an already profound

207 methodological gap in sequencing approaches between molecular ecologists in the Global North
208 and the Global South (**Table 1**).

209 Our analysis contrasts patterns inferred from both a traditional source of biodiversity
210 data—vouchered specimens in NHCs—and a contemporary genomic resource archive, the
211 National Institute of Health's NCBI Genome Browser. Our description of the latitudinal gradient
212 of total tetrapod species richness is broadly similar to other recent macroecological studies of the
213 phenomenon (Roll et al. 2017; Quintero et al. 2023), with global hotspots concentrated in
214 northwest South America and Central Africa (**Figure 1C**) and an approximately monotonic
215 decline in latitudinal species richness maxima from the equator to the poles. In contrast, the
216 latitudinal gradient of richness of tetrapod species with reference genomes is flattened, showing
217 only moderate declines at high latitudes and a mid-latitude peak in species richness in the
218 Northern Hemisphere (**Figure 1B**).

219 This difference is especially notable because we made no effort to correct for disparities
220 in historical specimen collection across latitude, with the consequence that our 'true' species
221 richness gradient significantly underestimates biodiversity in the tropics. Across longitude, our
222 analysis appears to underestimate diversity in East Asia, Indonesia, and Oceania (Quintero et al.
223 2023), likely due to both the coarse grain of our study and the region's greater distance from the
224 large NHCs in Europe and North America that are the backbone of curated GBIF data.
225 Geopolitical factors complicate a simple interpretation of these results: use of GBIF and NCBI
226 by scientists outside the Anglosphere may be limited by data sovereignty laws or other impacts
227 of economic and military rivalries. Regardless, species with publicly available reference
228 genomes as of July 2024 are more reflective of socioeconomic conditions than biogeographic
229 reality.

230 If both natural history collections and contemporary bioinformatics resources reflect
231 historical inequalities in development and scientific capacity, why do data from the former better
232 approximate the latitudinal gradient of species richness? Part of the answer lies in their different
233 goals: while the mission of many NHCs is to explicitly catalog and archive regional or global
234 biodiversity, the NCBI Genome Browser is typically used as a repository for open data
235 publication requirements and is less often an end unto itself. Another possibility is that
236 researchers looking to obtain genetic material from the tropics face unique obstacles to their
237 work. As the birth and golden age of NHCs coincided with the heyday of Western colonialism
238 (De Vos 2007; Quintero Toro 2012), scientific collecting was less restricted by concerns of
239 either sovereignty or Indigenous land tenure, let alone local or international regulations on the
240 transfer of biological samples (Kahn & Tyagi 2021; Sherkow et al 2022). Relatedly, one early
241 reader of this manuscript pointed to the added challenge of obtaining high-quality tissue samples
242 in hot, humid conditions.
243 We do not believe access to sources of DNA is a major obstacle: Science is rich with
244 examples of researchers overcoming far greater hurdles to their work, and extant frozen tissue
245 collections at institutions in the both the Global South and Global North include thousands of
246 samples awaiting sequencing. Instead, we suggest that evolving norms and persistent obstacles of
247 cost have led many relatively well-resourced scientists in the Global North to prioritize
248 generating reference genomes for local taxa. In spite of rapid declines in the per base pair cost of
249 whole-genome sequencing (Lou et al. 2021), high-coverage sequencing remains a significant
250 expense: while averages are hard to come by in this increasingly privatized sector, one of us
251 recently paid ~$13,000 USD (September 2023) for high-depth long-read and HiC sequencing of
252 a North American passerine bird from an industry contractor. Even in North America, this figure

253 likely pushes small, single PI labs to prioritize investing in generating resources for species at the
254 center of their research program, or otherwise likely to provide long-term utility—which are
255 often those in their own backyards.

256 In the spirit of the traditional mission of NHCs, the past decade has seen several
257 interrelated, international initiatives to increase taxonomic diversity in high-quality reference
258 genomes (e.g. O'Brien et al. 2014; Koepfli et al. 2015; Cheng et al. 2018; Rhie et al. 2021). Of
259 these initiatives, the Earth BioGenome Project (EBP)—a "moonshot" endeavor to sequence all
260 Eukarya within ten years—is by design the most ambitious, bringing its smaller-scale
261 predecessors together under a common set of goals and standards (Lewin et al. 2018). These
262 collaborations have had a profound impact on biodiversity genomics: As of March 2021, EBP
263 estimated at least one species from 15.6% of all Eukaryotic families had been sequenced,
264 marking notable progress towards its Phase I goal of complete family-level representation within
265 3 years. Though this focus on phylogenetic diversity has introduced its own bias by
266 undersampling lineages with high diversification rates, the initiative has nonetheless helped to
267 close what was surely an even larger gap in empirical patterns of species richness and species
268 richness represented by NCBI. In some cases, EBP or the initiatives under its umbrella have also
269 provided researchers with early access to draft assemblies of nonmodel organisms (e.g., Linck et
270 al. 2020) or with support and resources to produce new assemblies (e.g., Cadena et al. 2024), a
271 scenario that suggests the resource gap may be slightly less dire in practice than reported here.

272 Yet in a world of limited time, finite resources, and incompletely described biodiversity,
273 sequencing at scale is not immune to its own biases. For example, a recent publication
274 introducing an attempt to generate reference genomes for all vertebrates (Rhie et al. 2021)
275 included 127 authors affiliated with 102 institutions. Of these, only 13 are in the Global South: 4

276 in China, 3 in Korea, 2 in Malaysia, 1 in Singapore, 1 in Qatar, and 1 in Colombia. No authors or
277 institutional affiliations from Africa, Indonesia, or Oceania outside of Australia and New
278 Zealand were included. To date, institutions in the Global North have generated most animal
279 (Hotaling et al. 2021) and plant (Marks et al. 2021) genomes available on NCBI's GenBank, a
280 situation mirrored by an analysis of squamates alone by Pinto et al. (2023). Though
281 understandable considering the current distribution of scientists and resources—though not of
282 species—the imbalance seems likely to perpetuate representational biases in the near-term.
283 Current efforts by international collaborations like the Amphibian Genomics Consortium
284 to increase geographic representation and to offer support and opportunities to researchers from
285 developing countries and underrepresented groups (Kosch et al. 2024) are steps in the right
286 direction in the path to make the field of biodiversity genomics more equitable. We are
287 particularly enthusiastic about EBP's model of pairing targeted sequencing efforts with capacity
288 building workshops (see Sharaf et al. 2024 for a recent example from 11 African countries). In
289 line with the Convention on Biological Diversity's Nagoya Protocol (Secretariat of the
290 Convention on Biodiversity 2011), another critical dimension of the conversation about equitable
291 generation of encyclopedias of reference genomes is the need for researchers to build strong
292 partnerships with Indigenous peoples and other local communities, allowing them to participate
293 in and benefit from the different phases and products of sequencing projects (Ambler et al. 2020;
294 Colella et al. 2023; Mc Cartney et al. 2023).
295 If tropical species are underrepresented on NCBI, we would expect that they are only
296 rarely studied using whole-genome resequencing (and other sequencing strategies dependent on
297 a reference genome). Our review of conservation genetics papers published in *Molecular*
298 *Ecology, Journal of Heredity, Conservation Genetics,* and *Conservation Biology* over the last

299 five years suggests this is indeed the case. While the Global South / Global North binary and
300 measures of human development more generally are only imperfectly correlated with latitude,
301 their association nonetheless indicates that WGS is only rarely applied in conservation genetics
302 studies in the tropics (n=6), and almost never by a leading researcher with a primary affiliation to
303 a research institution in the region (n=1) (**Table 1**).

304 Similarly, for Global South scientists and for Global South focal taxa, microsatellites
305 remain the most common molecular approach; for scientists and focal taxa in the Global North,
306 reduced representation approaches dominate. We believe this is reflective of the expense and
307 limited availability of high-throughput sequencing in the Global South, regardless of whether
308 reads are assembled *de novo* or aligned to a reference. Lastly, we point out that across all
309 categories, the vast majority of scientists (87%) and species (77%) in our sample originate in the
310 Global North. Though our choice of North American or European-based journals precludes
311 generalization, the primacy of the English language in global science (Montgomery 2013) leads
312 us to interpret this as indicating the field of conservation genetics—let alone conservation
313 genomics—is in its infancy in the tropics. We thus doubt the patterns described here would
314 change appreciably if we were to include publications in other languages. Setting appropriate
315 goals, targets, and indicators to effectively conserve and monitor global genetic diversity will
316 require this situation to be remedied (Hoban et al. 2021).

317 We highlight discrepancies between available reference genomes and global
318 biogeographic patterns to encourage increased, equitable collaboration between scientists in the
319 Global North and Global South (see Sawchuk et al. 2024 for an African-based perspective on
320 equity and inclusion in population genetics and ancient DNA studies). In light of this, we make
321 three simple recommendations (see also Bertola et al. 2023). First, we encourage scientists from

resource-rich institutions to consider allocating effort and funds towards generating reference genomes that serve the needs of managers and researchers in the Global South. Second, we support the continued development of multinational sequencing projects, but ask funders and senior personnel to increasingly consider prioritization, inclusion, and capacity building in areas of the world with rich biodiversity and limited resources to study it using genomic tools Third, we ask journals to consider issues of access and cost in editorial guidelines and decisions: While high-throughput sequencing is increasingly expected by editorial boards and reviewers at high-impact journals, it is not essential to address a variety of research questions (Bertola et al. 2023) and remains out of reach for most scientists residing where most of the world's species occur. If ecology, evolution, and conservation aim to accurately catalog and effectively protect life on Earth, remedying inequalities in genomic resources should be a major priority.

Data Accessibility: Processed GBIF datasets, NCBI metadata, and a digital notebook containing code to perform these analyses and generate Figure 1 are available at https://github.com/elinck/lat_grad_genome and from Data Dryad (https://doi.org/10.5061/dryad.2v6wwpzxh).

Acknowledgments: We thank Marty Kardos for the invitation to participate in this special issue. Carolyn Hogg and an anonymous reviewer provided helpful feedback on the study. E.B.L. expresses appreciation for his mobile phone's Wi-Fi hotspot, which was essential for submitting the initial draft of this manuscript from the field. Ongoing genomic work by C.D.C. is supported by the Facultad de Ciencias at Universidad de Los Andes through a *Programa de Investigación* (Convocatoria 2024-2026).

References:

Ambler, J., Dearden, P. K., Wilcox, P., Hudson, M., & Tiffin, N. (2021). Including digital sequence data in the Nagoya Protocol can promote data sharing. Trends in Biotechnology, 39(2), 116-125.

Asase, A., Mzumara-Gawa, T. I., Owino, J. O., Peterson, A. T., & Saupe, E. (2022). Replacing "parachute science" with "global science" in ecology and conservation biology. Conservation Science and Practice, 4(5), e517.

Avise, J. C. (2000). Phylogeography: the history and formation of species. Harvard University Press.

Bertola, L. D., Brüniche-Olsen, A., Kershaw, F., Russo, I. R. M., MacDonald, A. J., Sunnucks, P., ... & Segelbacher, G. (2024). A pragmatic approach for integrating molecular tools into biodiversity conservation. Conservation Science and Practice, 6(1), e13053.

Bradley, M. (2008). On the agenda: North–South research partnerships and agenda-setting processes. Development in Practice, 18(6), 673-685.

Cadena, C. D., Pabón, L., DoNascimiento, C., Abueg, L., Tilley, T., O-Toole, B., ... & Torres, M. (2024). A reference genome for the Andean cavefish Trichomycterus rosablanca (Siluriformes, Trichomycteridae): Building genomic resources to study evolution in cave environments. Journal of Heredity, 115(3), 311-316.

Colella, J. P., Silvestri, L., Súzan, G., Weksler, M., Cook, J. A., & Lessa, E. P. (2023). Engaging with the Nagoya Protocol on Access and Benefit-Sharing: recommendations for noncommercial biodiversity researchers. Journal of Mammalogy, 104(3), 430-443.

Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: addressing disparity in global monitoring. Tropical Conservation Science, 1(2), 75-88.

Cornwell, W. K., Pearse, W. D., Dalrymple, R. L., & Zanne, A. E. (2019). What we (don't) know about global plant diversity. Ecography, 42(11), 1819-1831.

Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., Ram, K. (2024). rgbif: Interface to the Global Biodiversity Information Facility API. R package version 3.8.0, https://CRAN.R-project.org/package=rgbif.

Cheng, S., Melkonian, M., Smith, S. A., Brockington, S., Archibald, J. M., Delaux, P. M., ... & Wong, G. K. S. (2018). 10KP: A phylodiverse genome sequencing plan. Gigascience, 7(3), giy013.

De Vos, P. S. (2007). Natural history and the pursuit of empire in eighteenth-century Spain. Eighteenth-Century Studies, 40(2), 209-239.

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity, 107(1), 1-15.

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. Trends in Ecology & Evolution, 29(1), 51-63.

Feeley, K. J., & Silman, M. R. (2011). The data void in modeling current and future distributions of tropical species. Global Change Biology, 17(1), 626-630.

Freeman, B. G., & Pennell, M. W. (2021). The latitudinal taxonomy gradient. Trends in Ecology & Evolution, 36(9), 778-786.

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. Molecular Ecology, 26(20), 5369-5406.

GBIF.org. (2024a). GBIF Occurrence Download https://doi.org/10.15468/dl.p4r6hf Accessed from R via rgbif (https://github.com/ropensci/rgbif) on 2024-06-12.

GBIF.org. (2024b). GBIF Occurrence Download https://doi.org/10.15468/dl.njrdnn Accessed from R via rgbif (https://github.com/ropensci/rgbif) on 2024-06-20.

GBIF.org. (2024c). GBIF Occurrence Download https://doi.org/10.15468/dl.zw99uq Accessed from R via rgbif (https://github.com/ropensci/rgbif) on 2024-06-20.

GBIF.org. (2024d). GBIF Occurrence Download https://doi.org/10.15468/dl.jbg6xy Accessed from R via rgbif (https://github.com/ropensci/rgbif) on 2024-06-21.

GBIF.org. (2024e). GBIF Occurrence Download https://doi.org/10.15468/dl.5pkrsd Accessed from R via rgbif (https://github.com/ropensci/rgbif) on 2024-06-21.

GBIF.org. (2024f). GBIF Occurrence Download https://doi.org/10.15468/dl.jd85b2 Accessed from R via rgbif (https://github.com/ropensci/rgbif) on 2024-07-05.

GBIF.org. (2024g). GBIF Occurrence Download https://doi.org/10.15468/dl.59eyey Accessed from R via rgbif (https://github.com/ropensci/rgbif) on 2024-07-05.

GBIF.org. (2024h). GBIF Occurrence Download https://doi.org/10.15468/dl.vybgce Accessed from R via rgbif (https://github.com/ropensci/rgbif) on 2024-07-05.

Hillis, D. M., Moritz, C., & Mable, B. K. (Eds.). (1996). Molecular Systematics (2nd ed.). Sinauer.

Hoban, S. M., Hauffe, H. C., Pérez-Espona, S., Arntzen, J. W., Bertorelle, G., Bryja, J., ... & Bruford, M. W. (2013). Bringing genetic diversity to the forefront of conservation policy and management. Conservation Genetics Resources, 5, 593-598.

Hoban, S., Bruford, M. W., Funk, W. C., Galbusera, P., Griffith, M. P., Grueber, C. E., ... & Vernesi, C. (2021). Global commitments to conserving and monitoring genetic diversity are now necessary and feasible. Bioscience, 71(9), 964-976.

Hogg, C. J. (2024). Translating genomic advances into biodiversity conservation. Nature Reviews Genetics, 25(5), 362-373.

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. Molecular Ecology, 30(1), 62-82.

Hotaling, S., Kelley, J. L., & Frandsen, P. B. (2021). Toward a genome sequence for every animal: Where are we now? Proceedings of the National Academy of Sciences, 118(52), e2109019118.

Hull, D. L. (1990). Science as a process: an evolutionary account of the social and conceptual development of science. University of Chicago Press.

Khan, A., & Tyagi, A. (2021). Considerations for initiating a wildlife genomics research project in South and South-East Asia. Journal of the Indian Institute of Science, 101(2), 243-256.

King, D. A. (2004). The scientific impact of nations. Nature, 430(6997), 311-316.

Koepfli, K. P., Paten, B., Genome 10K Community of Scientists, & O'Brien, S. J. (2015). The Genome 10K Project: a way forward. Annual Review of Animal Biosciences, 3(1), 57-111.

Kosch, T. A., Torres-Sanchez, M., Liedtke, H. C., Summers, K., Yun, M. H., Crawford, A. J., ... & Amphibian Genomics Consortium (AGC). (2024). The Amphibian Genomics Consortium: advancing genomic and genetic resources for amphibian research and conservation. bioRxiv, 2024-06.

Lemmon, E. M., & Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. Annual Review of Ecology, Evolution, and Systematics, 44, 99-121.

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., ... & Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. Proceedings of the National Academy of Sciences, 115(17), 4325-4333.

Lewin, H. A., Richards, S., Lieberman Aiden, E., Allende, M. L., Archibald, J. M., Bálint, M., ... & Zhang, G. (2022). The earth BioGenome project 2020: Starting the clock. Proceedings of the National Academy of Sciences, 119(4), e2115635118.

Linck, E., Freeman, B. G., & Dumbacher, J. P. (2020). Speciation and gene flow across an elevational gradient in New Guinea kingfishers. Journal of Evolutionary Biology, 33(11), 1643-1652.

Logino, H. (2019). The Social Dimensions of Scientific Knowledge. In E. N. Zalta (Ed.), The Stanford Encylopedia of Philosophy (Summer 2019 Edition). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2019/entries/scientific-knowledge-social/

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. Molecular Ecology, 30(23), 5966-5993.

Marks, R. A., Hotaling, S., Frandsen, P. B., & VanBuren, R. (2021). Representation and participation across 20 years of plant genome sequencing. Nature Plants, 7(12), 1571-1578.

Massicotte, P., South, A. (2024). rnaturalearth: World Map Data from Natural Earth. R package version 1.0.1.9000, https://github.com/ropensci/rnaturalearth, https://docs.ropensci.org/rnaturalearthhires/, https://docs.ropensci.org/rnaturalearth/.

May, R. M. (1997). The scientific wealth of nations. Science, 275(5301), 793-796.

Mc Cartney, A. M., Head, M. A., Tsosie, K. S., Sterner, B., Glass, J. R., Paez, S., ... & Hudson, M. (2023). Indigenous peoples and local communities as partners in the sequencing of global eukaryotic biodiversity. npj Biodiversity, 2(1), 8.

Melles, S. J., Scarpone, C., Julien, A., Robertson, J., Levieva, J. B., Carrier, C., ... & Morales, K. (2019). Diversity of practitioners publishing in five leading international journals of applied ecology and conservation biology, 1987–2015 relative to global biodiversity hotspots. Ecoscience, 26(4), 323-340.

Meneghini, R., Packer, A. L., & Nassi-Calo, L. (2008). Articles by Latin American authors in prestigious journals have fewer citations. PloS ONE, 3(11), e3804.

Montgomery, S. L. (2013). Does science need a global language?: English and the future of research. University of Chicago Press.

Nori, J., Loyola, R., & Villalobos, F. (2020). Priority areas for conservation of and research focused on terrestrial vertebrates. Conservation Biology, 34(5), 1281-1291.

OBrien, S. J., Haussler, D., & Ryder, O. (2014). The birds of Genome10K. GigaScience, 3(1), 2047-217X.

O'Leary, N. A., Cox, E., Holmes, J. B., Anderson, W. R., Falk, R., Hem, V., ... & Schneider, V. A. (2024). Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. Scientific Data, 11(1), 732.

Opitz, D. L. (2004). Aristocrats and professionals: Country-house science in late-Victorian Britain [Doctoral dissertation, University of Minnesota]. ProQuest Dissertations and Theses Global.

Opitz, D. L. (2006). " This House is a Temple of Research": Country-House Centres for Late-Victorian Science. In Cliffor, D., Wadge, E., Warwick, A., & Willis, M, (Eds.), Repositioning Victorian sciences: shifting centres in nineteenth-century thinking. Anthem Press.

Pasterkamp, G., Rotmans, J., de Kleijn, D., & Borst, C. (2007). Citation frequency: A biased measure of research impact significantly influenced by the geographical origin of research articles. Scientometrics, 70(1), 153-165.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal, 10(1), 439–446. doi:10.32614/RJ-2018-009.

Pebesma, E., Bivand, R. (2023). Spatial Data Science: With applications in R. Chapman and Hall/CRC. doi:10.1201/9780429459016.

Perez, T. M., & Hogan, J. A. (2018). The changing nature of collaboration in tropical ecology and conservation. Biotropica, 50(4), 563-567.

Pinto, B. J., Gamble, T., Smith, C. H., & Wilson, M. A. (2023). A lizard is never late: Squamate genomics as a recent catalyst for understanding sex chromosome and microchromosome evolution. Journal of Heredity, 114(5), 445-458.

Quintero, I., Landis, M. J., Jetz, W., & Morlon, H. (2023). The build-up of the present-day tropical diversity of tetrapods. Proceedings of the National Academy of Sciences, 120(20), e2220672120.

Quintero Toro, C. (2012). Birds of empire, birds of nation: A history of science, economy, and conservation in United States-Colombia relations. Ediciones Uniandes-Universidad de los Andes.

R Core Team. (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Reddy, S. (2014). What's missing from avian global diversification analyses? Molecular Phylogenetics and Evolution, 77, 159-165.

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., ... & Jarvis,

502 E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species.
503 Nature, 592(7856), 737-746.
504 Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A. M., Bernard, R., ... & Meiri,
505 S. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation.
506 Nature Ecology & Evolution, 1(11), 1677-1682.
507 Sawchuk, E. A., Sirak, K. A., Manthi, F. K., Ndiema, E. K., Ogola, C. A., Prendergast,
508 M. E., ... & Yimer, G. (2024). Charting a landmark-driven path forward for population genetics
509 and ancient DNA research in Africa. The American Journal of Human Genetics, 111(7), 1243-
510 1251.
511 Secretariat of the Convention on Biodiversity. (2011). Nagoya Protocol on Access to
512 Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization
513 to the Convention on Biological Diversity: text and annex. Montreal, Canada, Secretariat of the
514 Convention on Biodiversity, 15pp. doi: http://dx.doi.org/10.25607/OBP-789
515 Seth, S. (2009). Putting knowledge in its place: science, colonialism, and the
516 postcolonial. Postcolonial Studies, 12(4), 373-388.
517 Sharaf, A., Nesengani, L. T., Hayah, I., Kuja, J. O., Mdyogolo, S., Omotoriogun, T. C., ...
518 & Ebenezer, T. E. (2024). Establishing African genomics and bioinformatics programs through
519 annual regional workshops. Nature Genetics, 1-10.
520 Sherkow, J. S., Barker, K. B., Braverman, I., Cook-Deegan, R., Durbin, R., Easter, C. L.,
521 ... & Greely, H. T. (2022). Ethical, legal, and social issues in the Earth BioGenome Project.
522 Proceedings of the National Academy of Sciences, 119(4), e2115859119.
523 Soares, L., Cockle, K. L., Ruelas Inzunza, E., Ibarra, J. T., Miño, C. I., Zuluaga, S., ... &
524 Martins, P. V. R. (2023). Neotropical ornithology: Reckoning with historical assumptions,

removing systemic barriers, and reimagining the future. Ornithological Applications, 125(1), duac046.

Stocks, G., Seales, L., Paniagua, F., Maehr, E., & Bruna, E. M. (2008). The geographical and institutional distribution of ecological research in the tropics. Biotropica, 40(4), 397-404.

Stephan, P. (2012). How economics shapes science. Harvard University Press.

Taylor, H. R., Dussex, N., & van Heezik, Y. (2017). Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. Global Ecology and Conservation, 10, 231-242.

Titley, M. A., Snaddon, J. L., & Turner, E. C. (2017). Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. PloS ONE, 12(12), e0189577.

Toomey, A. H., Knight, A. T., & Barlow, J. (2017). Navigating the space between research and implementation in conservation. Conservation Letters, 10(5), 619-625.

Toews, D. P., Campagna, L., Taylor, S. A., Balakrishnan, C. N., Baldassarre, D. T., Deane-Coe, P. E., ... & Winger, B. M. (2016). Genomic approaches to understanding population divergence and speciation in birds. The Auk, 133(1), 13-30.

Willig, M. R., Kaufman, D. M., & Stevens, R. D. (2003). Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. Annual Review of Ecology, Evolution, and Systematics, 34(1), 273-309.

Vamosi, J. C., & Vamosi, S. M. (2008). Extinction risk escalates in the tropics. PLoS ONE, 3(12), e3886.

548 **Figures & Tables:**

549 **Table 1.** Summary of regional authorship affiliation, sequencing strategy, and sampled
550 focal species range for empirical conservation genetics papers from 2019-2024 in four leading
551 journals. Integers indicate the total number of studies in each category, while numbers in
552 parentheses refer to its proportion out of all reviewed articles (n=394). Regional authorship
553 affiliation is multually exclusive, while the distribution of focal taxa is not, as studies could
554 involve multiple taxa, broadly distributed taxa, or long-distance migrants. Papers were assigned
555 to a sequencing strategy based on the most data-intensive approach they employed (i.e., a study
556 applying both Sanger sequencing and microsatellites would be assigned to the 'Microsatellites'
557 category.)

558

Sequencing Strategy	**Global South Author**	**Only Global North Authors**	**Global South Taxon**	**Global North Taxon**
Sanger	22 (0.0558)	25 (0.0634)	26 (0.0660)	26 (0.0659)
Microsatellites	43 (0.1066)	105 (0.2690)	45 (0.1142)	108 (0.2741)
Reduced Representation	29 (0.0736)	129 (0.3299)	31 (0.0787)	131 (0.3325)
WGS	5 (0.0013)	24 (0.0609)	11 (0.0152)	18 (0.0279)
Other	4 (0.0101)	8 (0.0203)	4 (0.0102)	8 (0.0203)

559

560

561

562 **Table 2.** Summary of regional authorship affiliation, study goals, and sampled focal
563 species range among reviewed papers. Study goals refer to broad tiers of research questions with
564 increasing data requirements. Papers were assigned to each on the basis of the most data-
565 intensive analysis they employed (i.e., a paper inferring population genetic structure and
566 identifying loci under selection would be assigned to the tier 3). As in Table 1, 'Global South
567 Author' and 'Only Global North Authors' are mutually exclusive categories, while 'Global
568 South Taxon' and 'Global North Taxon' are not.
569

Study Goals	**Global South Author**	**Only Global North Authors**	**Global South Taxon**	**Global North Taxon**
1. Taxonomy / systematics, identification, or sexing	9 (0.0228)	19 (0.0635)	11 (0.0279)	17 (0.04315)
2. Phylogeography / population genetic structure, estimating genetic diversity, and inferring demographic history	90 (0.2284)	241 (0.6192)	102 (0.2589)	248 (0.6294)
3. Detecting outlier loci, quantifying runs of homozygosity, and evaluating adaptive potential	4 (0.0106)	31 (0.0786)	4 (0.0102)	31 (0.0787)

570
571

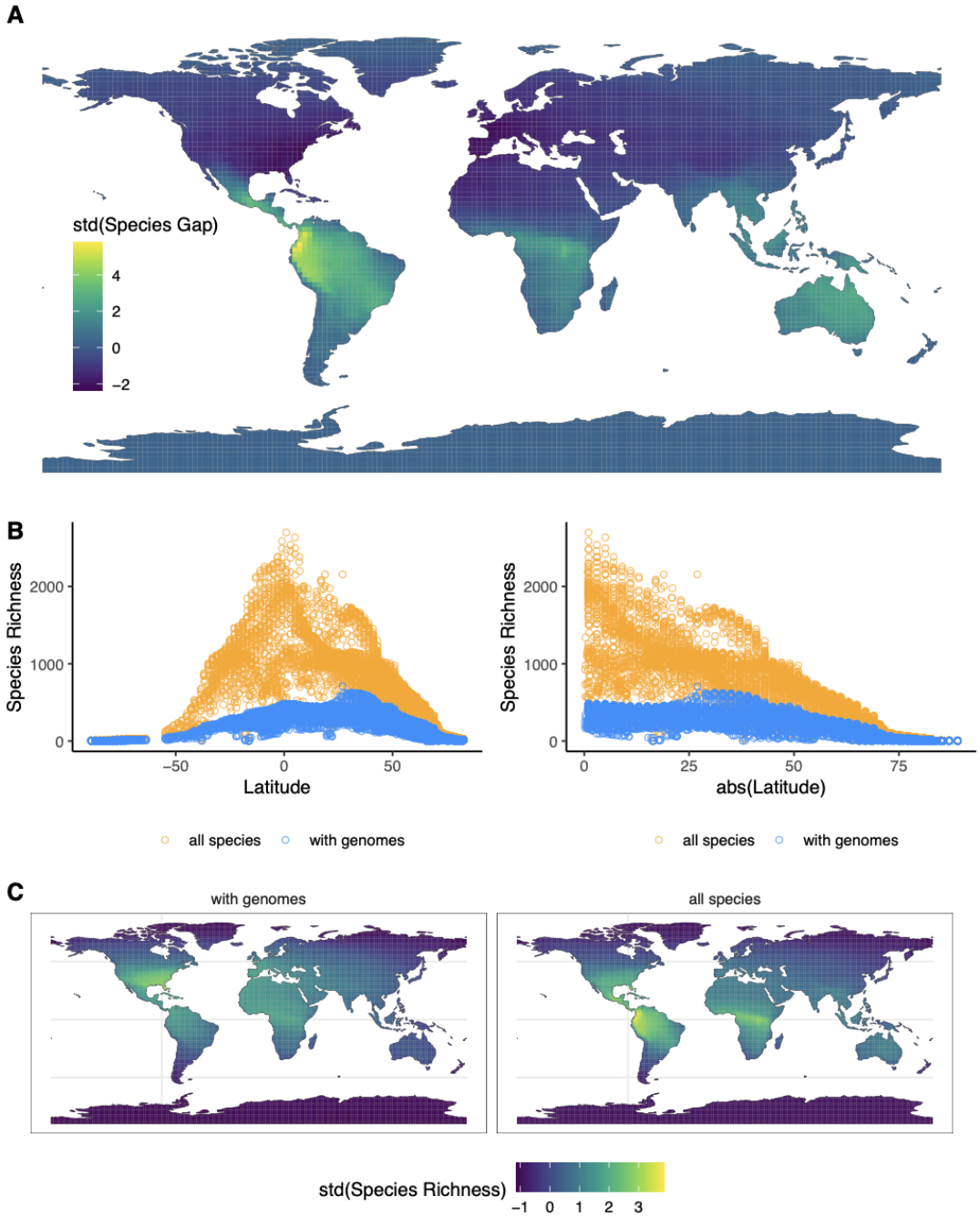

Figure 1. A) Tropical species are underrepresented among available reference genomes. Colors reflect the standardized difference between total species richness and the number of species with an assembled reference genome on the NCBI genome browser. Negative values indicate grid cells falling below calculated global mean species richness. B) Richness among species with reference genomes does not reflect global patterns of biodiversity. Blue circles represent total local species richness in 2-degree by 2-degree grid cells, while gold circles represent richness of species with assembled reference genomes. C) Global patterns of species richness calculated from species with reference genomes and all species with NHC specimen records on GBIF, respectively.